SI Appendix for

Regional control of histone H3 lysine 27 methylation in *Neurospora*

Kirsty Jamieson*, Michael R. Rountree*, Zachary A. Lewis, Jason E. Stajich and Eric U. Selker†

*These authors contributed equally to this work

†Corresponding Author: Eric Selker (email: selker@uoregon.edu)

SI Materials and Methods

cDNA preparation mRNA was purified from 100 μg of DNase-treated total RNA using the Oligotex mRNA Mini Kit (Qiagen Cat. # 70022). First and second strand cDNA synthesis was then conducted on 100 ng of mRNA with random primers using the SuperScript® Double-Stranded cDNA Synthesis Kit (Invitrogen Cat. # 11917-010). The cDNA was fragmented by sonication with a Branson Sonifier 450 on output 1.2, duty cycle 80, for 80 pulses. The fragmented cDNA was purified on Qiagen MinElute columns (Cat. #28206).

Preparation of ChIP-enriched DNA and double-stranded cDNA for sequencing ChIP-enriched DNA and double-stranded cDNA were prepared for sequencing by first blunting the ends of the fragments using the Quick Blunting Kit (New England Biolabs #E1201L) followed by purification of the DNA on a Qiagen MinElute column using the PCR purification protocol. Adenosines (A's) were then added to the blunted ends using Klenow fragment (3' to 5' exo-) (New England Biolabs #M0212S) followed by purification of the DNA on a Qiagen MinElute column using the reaction cleanup protocol. Paired-end adapters containing 6 bp barcodes, to allow for multiplexing, were ligated to the DNA fragments with Quick DNA Ligase (New England Biolabs #M2200S). The DNA was size-fractionated on a 2% agarose gel. "Invisible" fragments between 250-400 bp were excised and purified with the MinElute Qiagen Gel Purification system (Qiagen Cat. # 28604). Purified DNA was then amplified by PCR using the PfuTurbo Cx Hotstart DNA polymerase (Agilent Technologies Cat. #600410) using a limited number of cycles. Amplified DNA was size-fractionated on a 2% agarose gel in 1xTAE buffer. The "smear" of amplified DNA between 300-450 bp was excised from the gel and purified as before with the Qiagen MinElute Gel Purification kit. Purified DNA was quantitated using a Qubit fluorometer (Life Technologies), multiplexed with other samples containing different barcodes, and sequenced on either Illumina's Genome Analyzer II (40-nt read length) or HiSeq 2000 (50-nt read length) next-generation sequencers (Genomics Core Facility, University of Oregon).

Sequence analysis Sequence alignments were performed using Bowtie with default settings (1) and output in SAM format (2). The SAMTools tool kit was utilized to convert from the

SAM format to the BAM format and to remove PCR artifacts (rmdup tool) that appeared as large spikes in the data (2). RNA-Seq reads were mapped to the *N. crassa* OR74A reference genome using the default settings on TopHat (3). The mapped RNA-Seq reads were then used to estimate transcript abundance (FPKM - fragments per kilobase of exon per million fragments mapped) using the Cufflinks program (3). The Cuffdiff program was used to compare the relative abundance of transcripts between different RNA-Seq samples (3). For display purposes, ChIP-Seq and RNA-Seq reads were processed using the "count" function of igvtools (http://www.broadinstitute.org/igv/igvtools) to generate a tiled data file (.tdf) representing read densities over 300 bp windows across the genome. These .tdf files were displayed using the Integrative Genomics Viewer (IGV:http://www.broadinstitute.org/igv) (4).

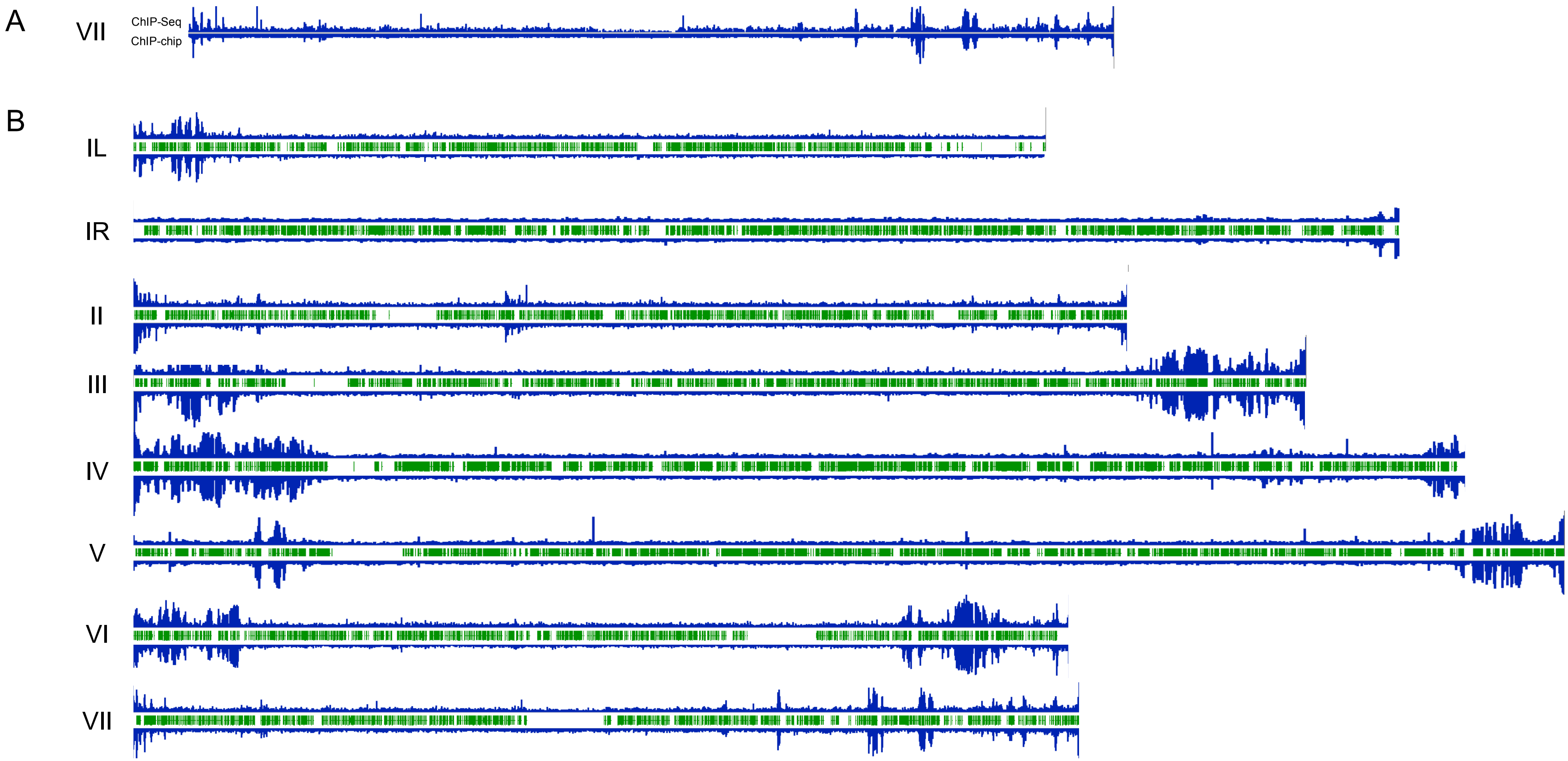
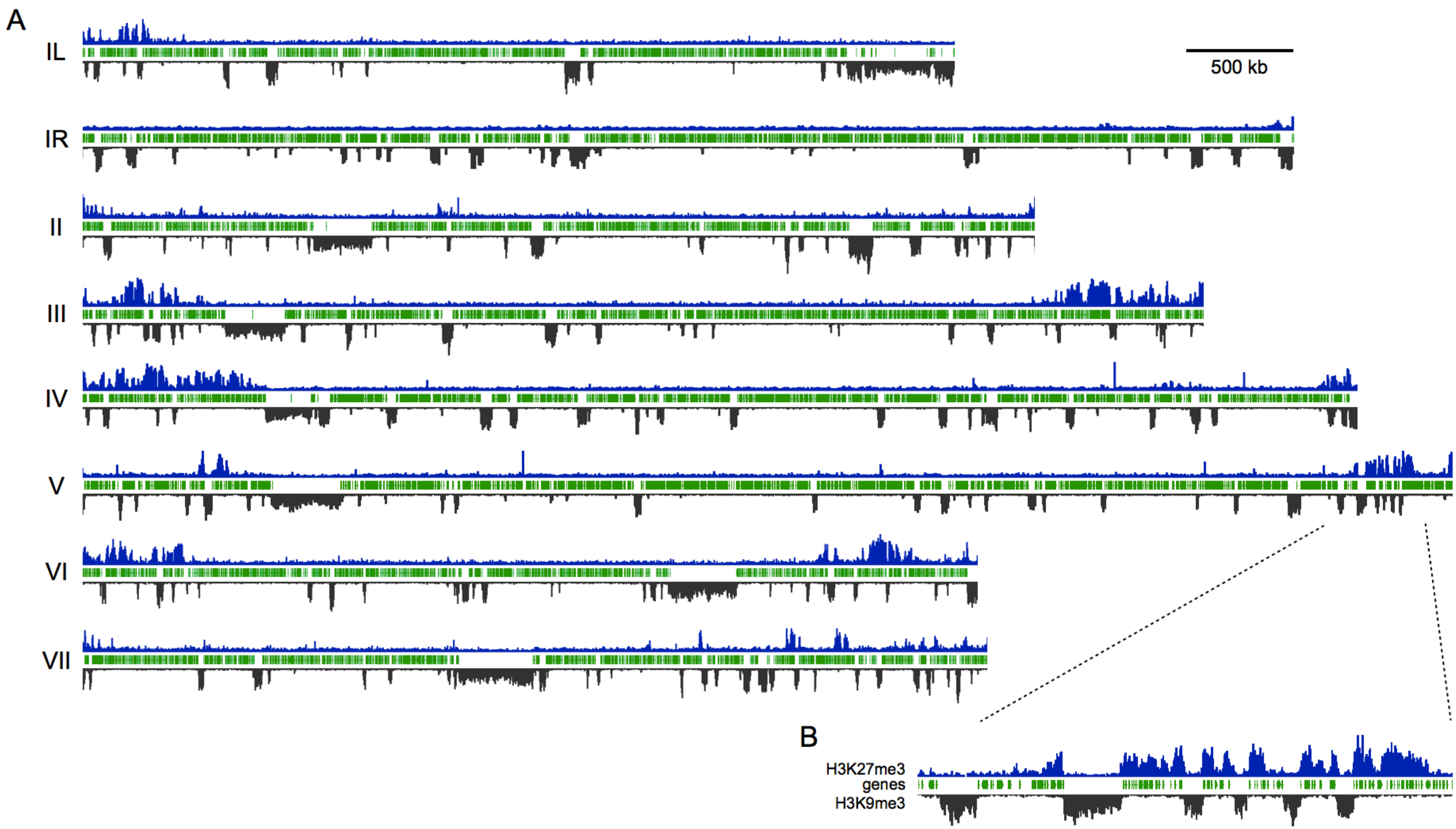

A
VII
ChIP-Seq
ChIP-chip
B
IL
IR
II
III
IV
V
VI
VII

Fig. S1. H3K27me3 domains are reproducibly detected. (A) The H3K27me3 ChIP-Seq read density profile (upper track) closely matches that obtained by ChIP-chip for LG VII (lower track). (B) H3K27me3 ChIP-Seq read density is shown for the seven LGs of *N. crassa* for cultures grown in Bird's medium (blue above the genes) and in Vogel's medium (inverted, blue below the genes). Virtually identical profiles were obtained.



A
IL
IR
II
III
IV
V
VI
VII
500 kb
B
H3K27me3
genes
H3K9me3

Fig. S2. Genome-wide distribution of H3K27me3 and H3K9me3 in *N. crassa*. (A) Predicted genes (vertical green lines), distribution of H3K27me3 (blue traces above genes) and H3K9me3 (black traces below genes) are represented to scale on the seven LGs of *N. crassa*. The largest chromosome, LG I, is divided at the right end of its centromere into IL and IR. (B) A portion of the right arm of LG V near the telomere is expanded to detail mutually exclusive H3K27me3 and H3K9me3 domains.

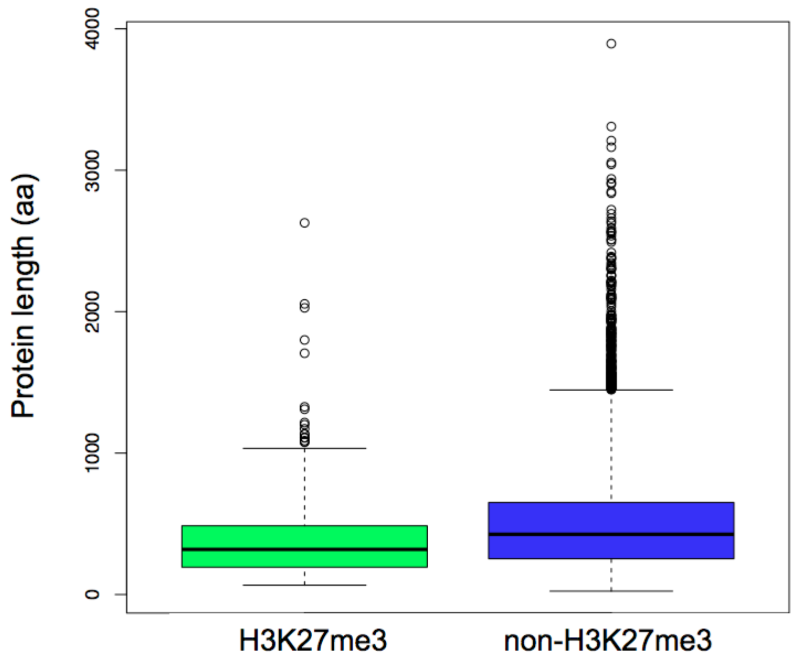


Fig. S3. Predicted lengths of proteins encoded by H3K27me3-marked and unmarked genes of *N. crassa*. Box-and-whisker plot of protein lengths show a significantly shorter average length for H3K27me3 marked genes ($P<2\times10^{-16}$, t-test).

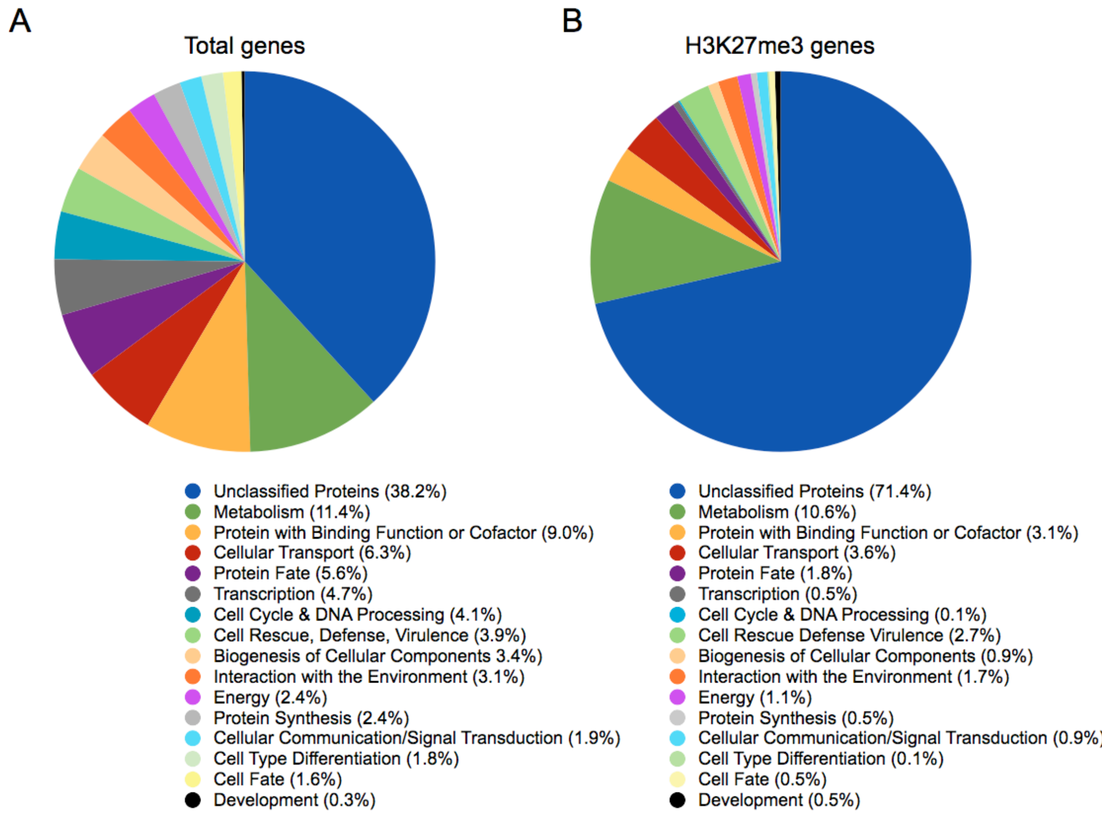


Fig. S4. Functional Category (FunCat) classification of *N. crassa* H3K27me3 genes. (A) Pie chart displaying the FunCat classification of all *N. crassa* predicted genes. (B) Pie chart displaying the FunCat classification of *N. crassa* genes found within H3K27me3 domains.

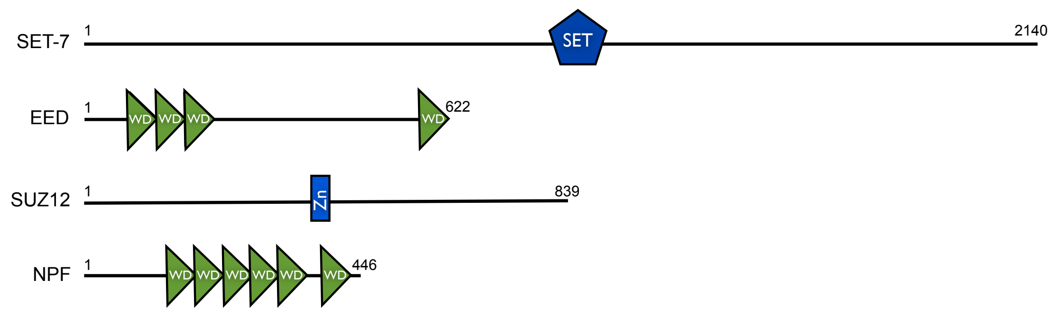
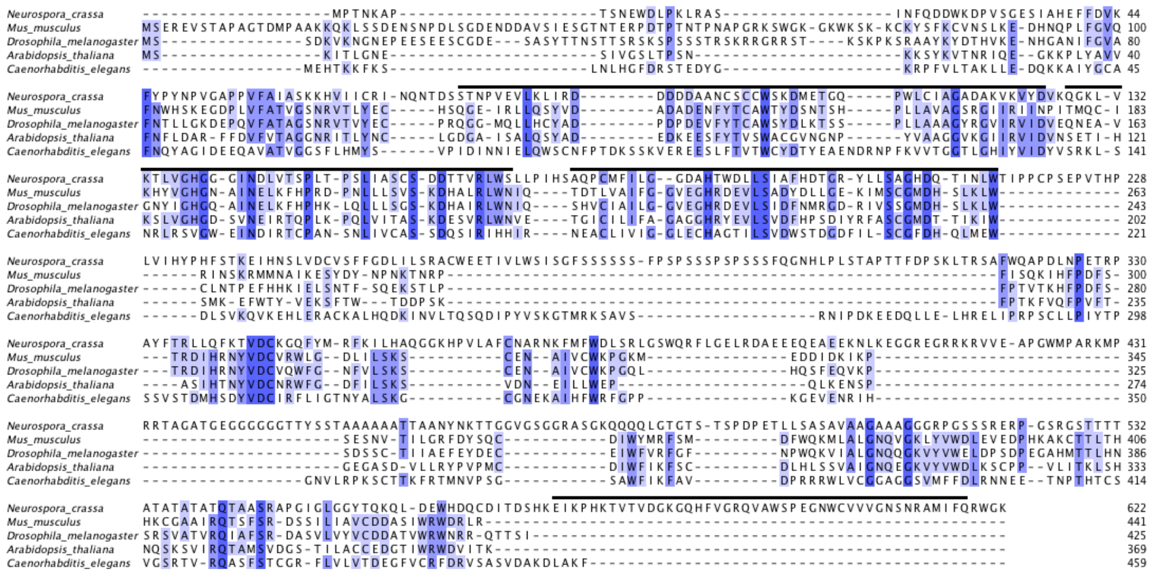


Fig. S5. Domain structures of the four core subunits of the *N. crassa* PRC2 complex. SET-7, the closest *Neurospora* homologue of *Drosophila* E(Z), contains the presumptive histone H3K27 methyltransferase domain (SET; blue pentagon); EED and NPF each possess multiple WD40 repeat domains (green triangles); *N. crassa* SU(Z)12 has a presumptive zinc-binding domain (blue rectangle).

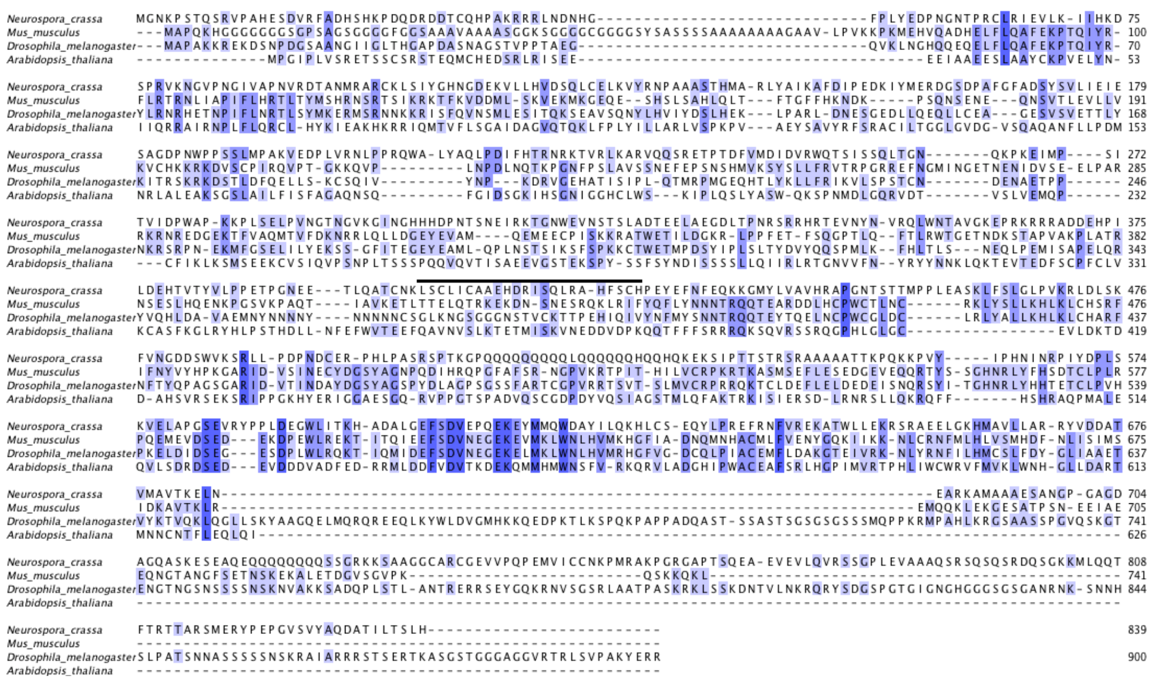
A


Neurospora_crassa
Mus_musculus
Drosophila_melanogaster
Arabidopsis_thaliana
Caenorhabditis_elegans

B


Neurospora_crassa
Mus_musculus
Drosophila_melanogaster
Arabidopsis_thaliana
Caenorhabditis_elegans

C


Neurospora_crassa
Mus_musculus
Drosophila_melanogaster
Arabidopsis_thaliana

D

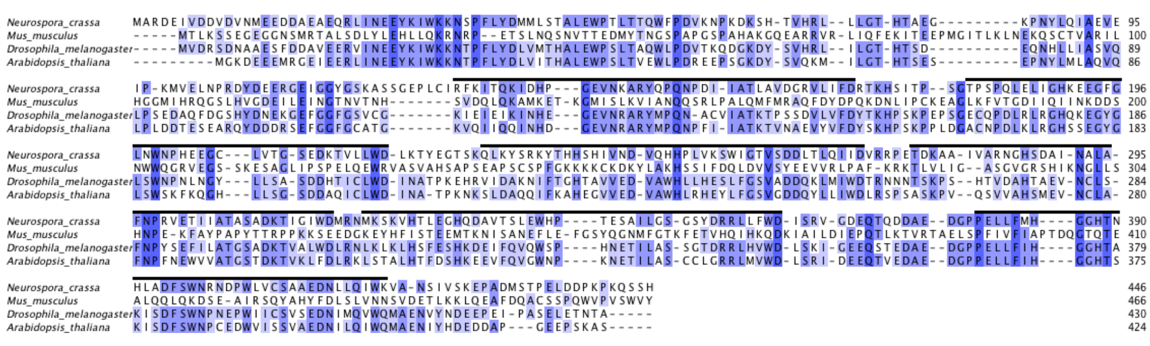

Neurospora_crassa
Mus_musculus
Drosophila_melanogaster
Arabidopsis_thaliana

Fig. S6. Multiple alignments of the PRC2 subunits. The protein sequence from each of the *N. crassa* PRC2 subunits was aligned to corresponding homologues from *Mus musculus*, *Drosophila melanogaster*, *Arabidopsis thaliana*, and *Caenorhabditis elegans* using ClustalWS. (A) Alignment of *N. crassa* SET-7 with homologs. ClustalWS alignments from *N. crassa* (XP_965043), *M. musculus* (AAH16391), *D. melanogaster* (NP_524021), *A. thaliana* (AEC07449) and *C. elegans* (O17514). The SET domain (V1040 to L1173, E = 3.38e-29) is indicated by the black line. (B) Alignment of *N. crassa* EED with homologs. ClustalWS alignments from *N. crassa* (XP_962071), *M. musculus* (NP_068676), *D. melanogaster* (AAA86427), *A. thaliana* (AEE76418) and *C. elegans* (Q9GYS1). Four WD40 domains are indicated by the black lines (S77 to D125, E = 7.96e+00; Q128 to S168, E = 1.74e-08; A175 to T216, E = 1.01e-04; E575 to Q618, E = 2.38e+01). (C) Alignment of *N. crassa* SUZ12 with homologs. ClustalWS alignments from *N. crassa* (XP_963451), *M. musculus* (AAH64461), *D. melanogaster* (Q9NJG9) and *A. thaliana* (AED96057). The zinc finger domain is indicated by the black line (L403 to H425, E = 2.12e+01). (D) Alignment of *N. crassa* NPF with homologs. ClustalWS alignments from *N. crassa* (XP_960994), *M. musculus* (AAC52970), *D. melanogaster* (AAF55146) and *A. thaliana* (AED97021). Five WD40 domains are indicated by the black lines (R130 to D169, E = 3.21e-01; T180 to D220, E = 1.02e-05; Q230 to D270, E = 9.94e-01; T276 to D317, E = 5.86e-06; M321 to D361, E = 2.14e-08; P378 to K418, E = 4.48e-02).

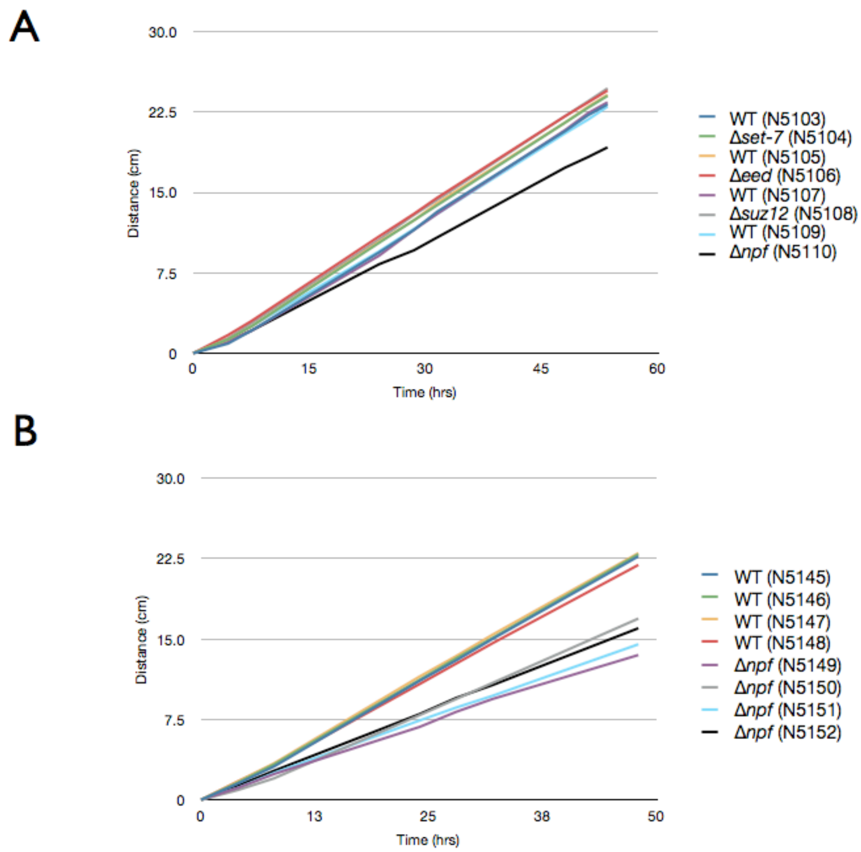


Fig. S7. Linear growth rates of PRC2 subunit deletion mutants. (A) The linear growth rate for wild-type and PRC2 deletion mutant strains (Table S5) was measured by growth in race tubes on Vogel's solid medium (5). (B) The linear growth rate was measured for four wild-type and four Δ*npf* strains (Table S5) on Vogel's solid medium (5).

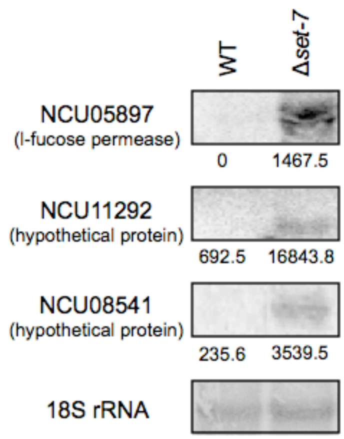


Fig. S8. Increased expression of H3K27me3 genes in the Δ*set-7* mutant. Northern blots show the increased expression of three additional genes (NCU05897, NCU11292 and NCU08541) in the Δ*set-7* strain. 18S rRNA stained with methylene blue is shown as a loading control.

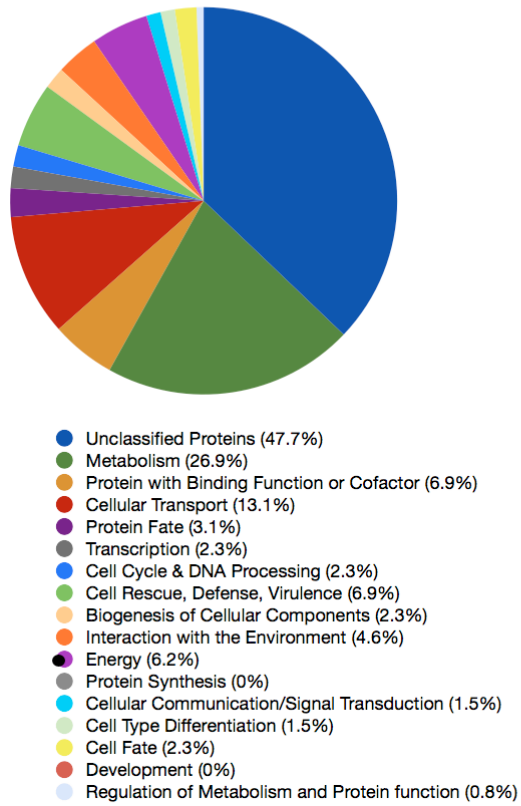


Fig. S9. Functional Category (FunCat) classification of genes showing increased expression of 130 upregulated genes in the Δ*set-7* strain.

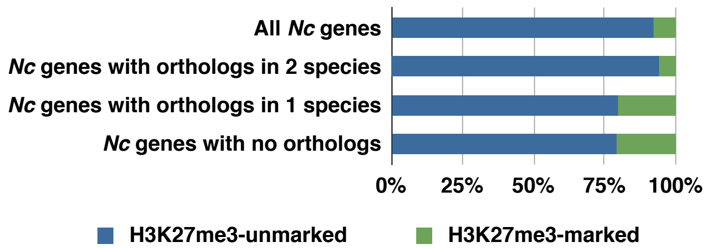


Fig. S10. Proportion of *N. crassa* H3K27me3-marked and –unmarked genes relative to their conservation in two other *Neurospora* species, *N. tetrasperma* (*N.t.*) and *N. discreta* (*N.d.*).

Table S1. H3K27me3 domain analysis summary in *N. crassa*

	H3K27me3 domains	Avg. Domain Size (kb)	Largest Domain (kb)	Total Genome Coverage (MB)	Genome Coverage (%)
WT (Bird's)	223	12.5	107.0*	2.8	6.8
WT (Vogel's)	232	12.1	107.5*	2.8	6.8
Δ*npf* (Vogel's)	187	8.4	79.5*	1.6	3.9

*The same domain on LG III. RSEG analysis carried out using a bin-size = 500 bp

Table S2. H3K27me3 domain analysis in *Neurospora* species

Species	H3K27me3 domains	Avg. Domain Size (kb)	Largest Domain (kb)	Total Genome Coverage (MB)	Genome Coverage (%)	H3K27me3-marked Genes
N. crassa	223	12.5	107.0	2.8	6.8	774
N. tetrasperma	167	10.8	101.5	1.8	4.6	536
N. discreta	186	14.1	94.5	2.6	7.1	822

RSEG analysis carried out using a bin-size = 500 bp

Table S3. H3K27me3 status of *N. crassa* orthologs in *N. discreta* and *N. tetrasperma*

N. crassa	*N. discreta*	*N. tetrasperma*	Gene Count
U	-	-	242
U	-	U	413
U	-	M	5
U	U	-	123
U	U	U	7927
U	U	M	11
U	M	-	8
U	M	U	186
U	M	M	21
M	-	-	66
M	-	U	32
M	-	M	78
M	U	-	8
M	U	U	85
M	U	M	44
M	M	-	25
M	M	U	136
M	M	M	258

Table S4. Purification of EED. An *N. crassa* strain bearing 3X-FLAG tagged EED at the amino-terminus and expressed under the *qa-2* promoter was used to purify the PRC2 complex. Associated proteins were identified by mass-spectrometry and the percent coverage of PRC2 members is indicated.

	MW (kDa)	Coverage (%)
SET-7 NCU07496	175.78	9
EED NCU05300	67.23	66.4
SUZ12 NCU05460	93.33	17.6
NPF NCU06679	50.41	74

Table S5. List of Strains

Experiment	Strain #	FGSC #	Species	Genotype
ChIP-Seq, qChIP, RNA-Seq, Northern blot	N3752	2489	*N. crassa*	*mat A*
ChIP- Seq, qChIP, RNA-Seq, Northern blot	N4718	11182	*N. crassa*	*mat a*; Δ*set-7::hph*
qChIP	N4719	14852	*N. crassa*	*mat A*; Δ*eed::hph*
qChIP	N4720	12769	*N. crassa*	*mat a*; Δ*su(z)12::hph*
ChIP- Seq, qChIP	N4721	13915	*N. crassa*	*mat a*; Δ*npf::hph*
ChIP- Seq	N5012	2508	*N. tetrasperma*	*mat A*
ChIP- Seq	N5014	8579	*N. discreta*	*mat A*
Race tubes	N5103	---	*N. crassa*	*mat A*
Race tubes	N5104	---	*N. crassa*	*mat A*; Δ*set-7::hph*
Race tubes	N5105	---	*N. crassa*	*mat A*
Race tubes	N5106	---	*N. crassa*	*mat a*; Δ*eed::hph*
Race tubes	N5107	---	*N. crassa*	*mat a*
Race tubes	N5108	---	*N. crassa*	*mat a*; Δ*su(z)12::hph*
Race tubes	N5109	---	*N. crassa*	*mat a*
Race tubes	N5110	---	*N. crassa*	*mat A*; Δ*npf::hph*
Race tubes	N5145	---	*N. crassa*	*mat A*
Race tubes	N5146	---	*N. crassa*	*mat a*
Race tubes	N5147	---	*N. crassa*	*mat A*
Race tubes	N5148	---	*N. crassa*	*mat A*
Race tubes	N5149	---	*N. crassa*	*mat A*; Δ*npf::hph*
Race tubes	N5150	---	*N. crassa*	*mat A*; Δ*npf::hph*
Race tubes	N5151	---	*N. crassa*	*mat a*; Δ*npf::hph*
Race tubes	N5152	---	*N. crassa*	*mat a*; Δ*npf::hph*

Table S6. List of primers

Experiment	Name	Sequence
qChIP	Gene 1 NCU06955 FP	GTCTTCGGGCATGGGTATAA
qChIP	Gene 1 NCU06955 RP	GATCAATCCTCTCGACTGGG
qChIP	Gene 2 NCU09590 FP	AGCATCCTCCACTGAGCACT
qChIP	Gene 2 NCU09590 RP	TCGAGTTTGGTAAGTGCTGTT
qChIP	Tel 1L NCU10129 FP	AGCGTTCAAATGCCGTGACCTGT
qChIP	Tel 1L NCU10129 RP	AGTCCAATGGTGCTAACGGCGA
qChIP	Tel 1R NCU10130 FP	GACGGACCTCTTCCGCTCGC
qChIP	Tel 1R NCU10130 RP	CCCTGCACGAGACGGTTCGA
qChIP	*hH4* NCU01634 FP	CATCAAGGGGTCATTCAC
qChIP	*hH4* NCU01634 RP	TTTGGAATCACCCTCCAG
Northern probe	NCU08907 FP	CTCACCACCCTCCTCGCCCTCGCC
Northern probe	NCU08907 RP	CCTCAAGCAGCACACAAATCCAAC
Northern probe	NCU05897 FP	CTATGGCCTCGGCGCCCTTCTCGCG
Northern probe	NCU05897 RP	GCCATTACAGGCCCTTCTCGCCGAC
Northern probe	NCU08541 FP	GCAATCAAAATGTCCGTCAACCGC
Northern probe	NCU08541 RP	GACTTGCAATGAGCCCTCAAGCC
Northern probe	NCU09663 FP	GTCGAGGCCGCCGCCTCCGTCTCC
Northern probe	NCU09663 RP	CTAGAAGAGACCAAGACCCATACC
Northern probe	NCU11292 FP	CGCTAGCAATATGGCAGGCAAACCG
Northern probe	NCU11292 RP	CCATCAACCTAAGCTTTCGATTCCC
HT-seq	PE-top adapter	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATC-barcode-T-3'
HT-seq	PE-bottom adapter	5'P-barcode-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
HT-seq	PE barcode #1	TAACCC (top adapter) / GGGTTA (bottom adapter)
HT-seq	PE barcode #2	TAAGGG (top adapter) / CCCTTA (bottom adapter)
HT-seq	PE barcode #3	TCAGTC (top adapter) / GACTGA (bottom adapter)
HT-seq	PE barcode #4	TCGCGC (top adapter) / GCGCGA (bottom adapter)
HT-seq	PE barcode #5	TCTTTCC (top adapter) / GGAAGA (bottom adapter)
HT-seq	PE barcode #6	TGCCGG (top adapter) / CCGGCA (bottom adapter)
HT-seq	PE barcode #7	TGTGTG (top adapter) / CACACA (bottom adapter)
HT-seq	PE barcode #8	TCCTTG (top adapter) / CAAGGA (bottom adapter)
HT-seq	PE barcode #9	TCACAG (top adapter) / CTGTGA (bottom adapter)
HT-seq	PE barcode #10	TGGTTC (top adapter) / GAACCA (bottom adapter)

List of SI Datasets

Dataset S1. H3K27me3 domains determined by RSEG (Bird's Medium)

Dataset S2. List of H3K27me3 genes (Bird's Medium)

Dataset S3. H3K27me3 domains determined by RSEG (Vogel's Medium)

Dataset S4. H3K27me3 domains in *N. tetrasperma* determined by RSEG

Dataset S5. H3K27me3 domains in *N. discreta* determined by RSEG

Dataset S6. H3K27me3 domains in the Δ*npf* strain determined by RSEG

Dataset S7. Differential domain analysis (RSEG-DIFF) comparing the Δ*npf* strain to wild-type

Dataset S8. List of genes upregulated in the Δ*set-7* strain (H3K27me3 genes in red and border genes in orange)

Dataset S9. List of genes down-regulated in wildtype vs. the Δ*set-7* strain

References

1. Langmead B (2010) Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics* Chapter 11:Unit 11.7.

2. Li H et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

3. Trapnell C et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28:511–515.

4. Robinson JT et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29:24–26.

5. Davis RH, De Serres FJ (1970) Genetic and microbiological research techniques for *Neurospora crassa*. *Meth Enzymol* 17A:47–143.